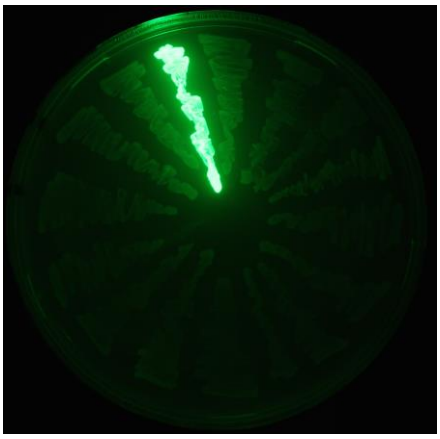


An analysis of the epistatic interactions of the *cis*- regulatory element in the arabinose operon of *Escherichia coli*

Anaísa Batista dos Santos Moreno
Mestrado em Biologia Celular e Molecular
Departamento de Biologia
2015

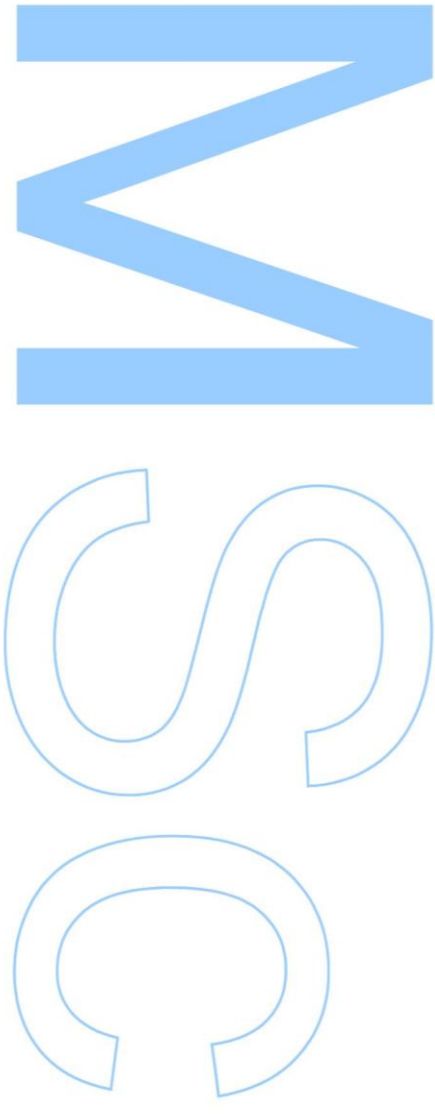
Orientador
Jonathan P. Bollback, Assistant Professor,
Institute of Science and Technology Austria

Coorientador
Mato Lagator, Postdoctoral Fellow,
Institute of Science and Technology Austria

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Acknowledgments

I would like to humbly acknowledge and thank all the inspiring people I worked with and helped me complete this project.

First and foremost I would like to thank Jonathan Bollback, without whom this would not be possible, for welcoming me into his group for a second time and for being a mentor in more than work.

I would like to acknowledge the people at IST Austria who made my stay amazing in so many ways. Mato Lagator thank you for all your help, patience and making me think more critically. Working with you and watching your work ethic helped me grow as a scientist. Hande Acar, thank you for always being there for me. Even in busy times, I knew I could always count on you for anything lab or life related. I also want to thank the other members of the Bollback group for all the helpful discussions and advice in the lab. My awesome office mate Tom Ellis, thank you for hearing my rants, reading my work and especially thank you for all the *hoppy* times. The rest of the Pho group – Chris Pull and Barbara Casillas Perez – you made life outside work so much fun, thank you.

I want to thank Professor Fernando Tavares for being my University coordinator, which allowed me to do the Erasmus+ internship. I also want to acknowledge the Erasmus+ mobility program, which supported me during my stay in Austria. My friends from Porto, Ana Amorim e Maria Costa, you girls are the best. Daniel, thank you for your friendship and indulging my love of everything post.

I would also like to thank those who contributed to life outside science and to maintaining my spirits high during the bad times: my family. My dad, Dominique, who unfortunately is no longer here to see me conclude this chapter of my life, thank you for always pushing, supporting and believing in me. Mom, thank you for your patience and for your continuous faith and support of my dreams. My amazing siblings, you are literally too many to thank by name, but I wouldn't be who I am without all of you, thank you for everything. Tina Almeida, you and your family are proof that there is more than one way to be a family, thank you for always welcoming me into your home. Finally, Patrik, there are many things to thank you for, so I will keep it simple: *tack älskling*.

Resumo

As bactérias são muito eficazes na regulação da expressão genética, pois este processo é crucial para o seu correto desenvolvimento. Alterações nos elementos reguladores podem ter consequências a níveis evolutivos, especialmente se estas mutações ocorrerem nas regiões de DNA em que os factores de transcrição se ligam: os elementos *cis*-reguladores. Apesar disto, pouco se sabe sobre os efeitos de mutações nestas regiões e da natureza das interações epistáticas entre as mesmas.

Neste estudo, um sistema sintético com os elementos reguladores do operão *araBAD* da bactéria *Escherichia coli*, importante no metabolismo de arabinose, foi utilizado para estudar as interacções epistáticas nesta região.

Para testar a existência de quaisquer interações epistáticas, mutações pontuais simples e duplas foram introduzidas em dois locais do operador, que controlam a expressão dos genes *araBAD*, necessários para o metabolismo da arabinose. Como as bactérias são expostas a uma variedade de ambientes, a natureza das interações epistáticas foi medida em dois ambientes diferentes.

Os resultados indicam que as interações epistáticas entre mutações neste elemento regulador são comuns e que na sua maioria são de uma natureza negativa. Também foi revelado que a magnitude da epistasia é dependente da localização das mutações nos operadores; e que existem interações entre o genótipo e o ambiente.

Palavras-chave: Epistasia; mutação; genótipo – ambiente; regulação da expressão genética; elementos reguladores da trancrição; operador; operão arabinose (*araBAD*).

Abstract

Bacteria are very effective at regulating gene expression, as this is crucial for the right development of organisms. Changes in the regulatory elements may have evolutionary consequences, particularly if these mutations occur in the DNA regions where transcription factors bind to: *cis*-regulatory elements. Despite this, little is known about the effects of mutations in these regions, and the nature epistatic interactions between these mutations.

In this study, a synthetic system with the *cis*-regulatory elements of the *araBAD* operon of the bacterium *Escherichia coli* was used to study the epistatic interactions in this region.

To test if there were any epistatic interactions, single and double point mutations were introduced in two operator sites, which control the expression of the *araBAD* genes, required for the metabolism of arabinose. As bacteria are often exposed to a multitude of environments, the nature of the epistatic interactions was measured in two different environments.

The results indicate that epistatic interactions between mutations in this regulatory element are common, and that they are mostly of a negative nature. It was also revealed that magnitude of epistatic interactions depended on the location of the mutations in the two operator sites; and that there are interactions between the genotype and environment.

Key words: Epistasis; mutation; genotype – environment; gene regulation; *cis*-regulatory elements; operator; arabinose operon *(araBAD)*.

Table of Contents

ACKNOWLEDGMENTS ..... **I**
RESUMO ..... **II**
ABSTRACT ..... **III**
TABLE OF CONTENTS ..... **IV**
LIST OF TABLES AND FIGURES ..... **V**
I. INTRODUCTION **1**
1. REGULATORY ELEMENTS 1
2. EPISTASIS 2
3. OPERONS 4
4. THE ARABINOSE OPERON 4
5. OBJECTIVES 9
II. MATERIALS AND METHODS **10**
1. BACTERIAL STRAINS, MEDIA AND GROWTH CONDITIONS 10
2. PLASMID CONSTRUCTION 10
3. CONSTRUCTION OF MUTANT LIBRARY 11
3.1 Site-directed Mutagenesis 11
3.2 Transformation of cells 11
3.3 Selection of transformants 12
3.4 Sequencing of mutants 12
3.5 Maintenance of mutant library 12
4. FITNESS ASSAYS 12
5. STATISTICAL ANALYSES 13
5.1 Expression Analysis 13
5.2 Estimation of Epistasis 13
III. RESULTS **15**
1. ASSEMBLY OF MUTANT LIBRARY 15
2. CHARACTERIZATION OF MUTATIONAL EFFECTS ON FITNESS 17
3. CHARACTERIZATION OF EPISTATIC INTERACTIONS 20
3.1 Epistatic interactions in different environments 20
3.2 Epistatic interactions relative to physical location of mutations 22
IV. DISCUSSION **23**
REFERENCES **28**
APPENDIX **33**

List of Tables and Figures

Table 1 – Reagents and volumes used in the PCR reactions 11
Table 2 – Single mutant library 16
Table 3 – Double mutant library 16

Figure 1 – Different types of epistasis along a path from a suboptimal allele *ab* towards the optimal *AB.* 3
Figure 2 – Genes required for the uptake and catabolism of L-arabinose in *E. coli.* 5
Figure 3 – AraC acts positively and negatively at p_{BAD} (±) and negatively at p_C (-) 5
Figure 4 – Proposed model for the regulation of the L-arabinose operon by arabinose. 6
Figure 5 – AraC binding to three adjacent major grooves of DNA. 7
Figure 6 – Sequences of the two direct repeats of *araI*. 7
Figure 7 – Missing contact data for AraC protein and *araI* operator. 8
Figure 8 – Architecture of the *araBAD* operon and its operators and of the synthetic construct . 15
Figure 9 – Relative fluorescence of single mutants in the presence and absence of arabinose (**A** and **B** respectively). 18
Figure 10 – Relative fluorescence of double mutants in the presence and absence of arabinose (**A** and **B** respectively). 19
Figure 11 – Network of mutant epistatic interactions. **A**. Epistatic interaction network in the presence of arabinose. **B**. Epistatic interaction network in the absence of arabinose. 21
Figure 12 – Mean epistasis of double mutants relative to their location on the operator half-sites in the presence (**A**) and absence (**B**) of arabinose. 22
Figure 13 – Possible binding outcomes caused my mutations in $araI_1$ and $araI_2$. 26

I. Introduction

1. Regulatory Elements

The right temporal and spatial control of gene regulation is crucial for the right development of organisms, core cellular processes and responses to environmental changes (Jeziorska *et al.*, 2009; Badis *et al.*, 2009). Deviances from the desired gene expression pattern may lead to many diseases, such as cancer (Levo and Segal, 2014).

In prokaryotes, gene expression is controlled by *cis*-regulatory DNA sequences, which encode the instructions for transcriptional activation and repression, and contain binding sites for *trans*-acting regulatory proteins (transcription factors) and other DNA binding proteins (Wittkopp, 2005; Jeziorska *et al.*, 2009; Kwasnieski *et al.*, 2012). The transcription factor-regulatory elements' interactions are responsible for regulation of gene expression in almost all biological processes (Tjian and Maniatis, 1994; Lelli *et al.*, 2012). *Cis*-regulatory elements can be promoters, enhancers, silencers, insulators and locus-control regions (LCRs) (Alexander *et al.*, 2010).

Due to the importance of correct gene regulation and their role in gene expression, *cis*-regulatory elements (CREs) have been a topic of many studies. Mutations in CREs often have unexpected effects on gene regulation, one reason for this is that CREs of catabolic genes and operons tend to be found either close to the promoter or in the downstream regions (Kwasnieski *et al.*, 2012; Sá-Nogueira *et al.*, 1997; Hueck and Hillen, 1995).

Changes in CREs represent an important part of the genetic basis for adaptation. Mutations in these regions may have a range of morphological, physiological and behavioral effects (Wray, 2007; Perez and Groisman, 2009; Wittkopp and Kalay, 2011). Studies of *cis*-regulatory sequence variance may help elucidate questions about the genetic mechanisms of phenotypic evolution and other fundamental biological phenomena (Wittkopp and Kalay, 2011; You and Yin, 2002).

2. Epistasis

Over 100 years ago William Bateson first used the term epistasis to describe the masking of the effects of one genetic variant by another (Bateson, 1909; Phillips, 2008). Over the years the term epistasis has had many different meanings depending on the different disciplines (Wade *et al.*, 2001; Hartl, 2014). Here epistasis will be described as any interaction between mutations that affects fitness or phenotype (Kussell, 2013).

Understanding epistasis is essential to understanding biology, and it is one of the major challenges of modern biology (Schenk *et al.*, 2013; Ibáñez-Marcelo *et al.* 2014). It is important to understand epistatic interactions between mutations as they may influence the course of evolution, in some cases hindering it by limiting the number of adaptive pathways; play a role in determining patterns of adaptation; and may explain how biological functions evolve (Schenk *et al.* 2013; de Vos *et al.* 2013[1]; Khan *et al.* 2011; Lehner, 2011).

Epistatic interactions can be classified into two categories: magnitude, sign and reciprocal epistasis. Magnitude epistasis means that the magnitude of the effect of a mutation is dependent on the presence of other mutations (**Figure** 1). Magnitude epistasis can be further classified into positive or synergistic epistasis, which describes epistatic interactions that result in a greater fitness effect than of that expected when mutations occur alone; and negative or antagonistic epistasis, which is a type of epistatic interaction between genes or mutants that have a smaller effect than expected from their effects when alone (Poelwijk *et al.* 2011; Lalic and Elena, 2012; Azevedo *et al.*, 2006). Magnitude epistasis is commonly observed in many different organisms and can influence the probability that a particular pathway is taken, thus contribution to the declining rates of adaptation observed in constant environments (Sanjuan and Elena, 2006; Schenk *et al.* 2013; Khan *et al.* 2011). Sign epistasis refers to cases where the fitness effect of a mutation can have a different sign depending on the genetic environment, i.e. it is possible for a mutation to be deleterious on a particular genetic background and beneficial on another (**Figure** 1) (Poelwijk *et al.,* 2011; Weinreich *et al*. 2005). Reciprocal sign epistasis is observed when sign of the fitness effect of a mutation is dependent on the state of another mutation (**Figure** 1) (Lalic and Elena, 20012; Poelwijk *et al.* 2011). Changes in the sign of epistasis have a strong impact on natural selection and may render certain mutational pathways inaccessible, thus enhancing the probability of other pathways, or even open up new pathways (Salverda *et al.* 2011; Schenk *et al.* 2013)

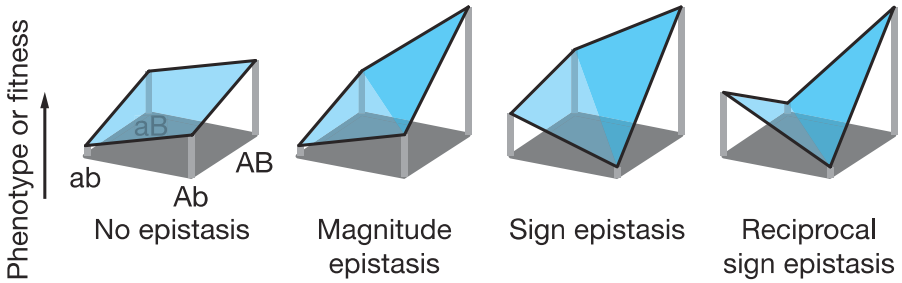


Figure 1 Different types of epistasis along a path from a suboptimal allele *ab* towards the optimal *AB* (Poelwijk *et al.* 2007)

Epistasis can be affected by many different factors, which can be difficult to understand and discern in complex organisms. It is known that the environment may affect fitness of certain mutations, so it is likely that it also has an effect on epistatic interactions (You and Yin, 2002; Claude *et al.*, 2014; de Vos *et al.*, 2013[2]).

Despite the importance of epistasis to understanding biological processes, little is known about the epistasis in complex organisms. Most studies range phages and viruses (Silander *et al.*, 2007; Parera *et al.*, 2007), bacteria (Maisnier-Patin *et al.*, 2005; Elena and Lenski, 1997) and yeast *Saccharomyces cerevisae* (Jasnos and Korona, 2007), and tend to focus on deleterious mutations (Schenk *et al.*, 2013; Sanjuán *et al.*, 2004). However as technologies to manipulate particular mutations; high-throughput functional genomics; and systems approaches to biology evolve and become more available, there is a renewed appreciation for the importance of studying epistasis, and for finding more unified and quantitative ways to answer these questions (Wang *et al.*, 2012; Phillips, 2008).

3. Operons

Monod *et al.* first used the term operon in 1960 to describe a set of genes transcribed by a single promoter (Jacob *et al.*, 1960; Schleif, 1993). Genes in an operon tend to be within physically proximity of each other and share related functions (Osbourn and Field, 2009). Operons contain different types of genetic elements: structural genes, these encode for the proteins required for the metabolic pathway or cellular process; a promoter, the starting point where RNA polymerase binds to start transcription, the promoter may be controlled by one or more regulatory proteins that respond to chemical signals or environmental stimuli; and the operator (may be more than one), normally found close to the promoter and acts as a control element to which a regulator protein binds, resulting in repression or activation of transcription (Ramos *et al.*, 2013). Operons are a common feature of bacterial and bacteriophages' genomes; it is estimated that 50% of the genes in *Escherichia coli* (*E. coli*) are transcribed as part of operons (Ralston, 2008; Lim *et al.*, 2011).

4. The Arabinose Operon

When its preferred source of carbon, glucose, is unavailable, *E. coli* is able to use L-arabinose, a pentose sugar, as a source of carbon and energy. For the metabolism of arabinose to take place, different genes are required to work jointly, two operons – *araBAD* and *araFGH*, and the genes – *araE* and *araC* (Schleif, 2010). AraFGH and AraE are responsible for the uptake of arabinose into the cell – transporters; *araBAD* codes for proteins that convert arabinose into a sugar easier to process by the cell; and *araC* acts as regulator (Schleif, 2000). The *araBAD* and its regulator, the AraC protein, have been the most studied (Brunelle and Schleif, 1989).

When arabinose enters the cell, the p_{BAD} promoter activates transcription of the *araBAD* operon, the protein enzymes coded by these genes start by converting L-arabinose to L-ribulose, by action of L-arabinose isomerase (AraA), in the next step the kinase encoded by *araB* phosphorylates L-ribulose to L-ribulose-5-phosphate and finally an epimerase (AraD) converts L-ribulose-5-phosphate to D-xylulose-5-phosphate, which in turn enters the pentose phosphate pathway to be further catabolized (**Figure** 2) (Gross and Englesberg, 1959; Englesberg, 1961; Schleif, 2000).

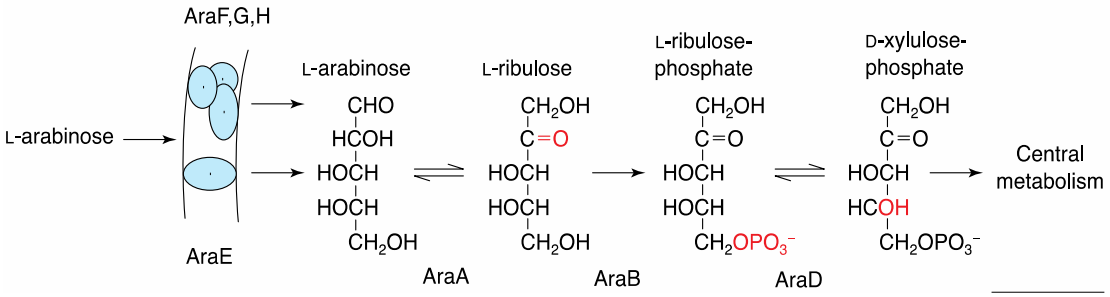


Figure 2 Genes required for the uptake and catabolism of L-arabinose in *E. coli* (Schleif 2000).

One of the factors that make this operon such an interesting subject is that it was the first operon that demonstrated a positive manner of regulation (Englesberg 1971, Englesberg and Wilcox 1974; Greenblatt and Schleif, 1971). Until then, due to previous experiments, most believed that operons were negatively regulated – the Operon Model (Stent 1964; Jacob *et al.* 1960; Schleif 1993). Although the exact mechanisms for regulation of the *araBAD* operon are still not fully understood, there is a proposed mechanism (Schleif, 1993; Niland *et al.*, 1996).

The arabinose operon is made up of the structural genes *araA*, *araB* and *araD*, their promoter p_{BAD}, the controlling sites *araI* and *araO* and the regulator gene *araC* (Gielow *et al.*, 1971). The *araI* site is made up of two contiguous regions – $araI_1$ and $araI_2$, and *araO* is also made up of two regions: $araO_1$ and $araO_2$; $araO_1$ can be further divided into two half-sites $araO_{1R}$ and $araO_{1L}$ (**Figure** 4) (Lee *et al.*, 1987; Schleif. 2003). The AraC protein both positively and negatively regulates the transcription of the *araBAD* operon (**Figure** 3) (Lee *et al.*, 1987; Englesberg, 1971).

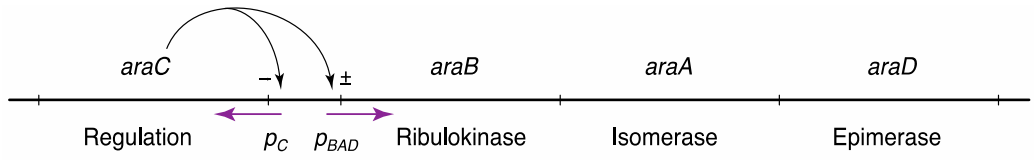


Figure 3 AraC acts positively and negatively at p_{BAD} (±) and negatively at p_C (-). (Schleif 2000)

In the regulation of the *araBAD* operon the AraC protein exists as homodimer, made of two monomers, which possesses a DNA-binding domain and a dimerization domain that has an arabinose-binding pocket and an N-terminal arm; the monomers are connected by a flexible linker (**Figure** 4 (a)) (Schleif, 2000).

In the absence of arabinose the two AraC monomers binds to two half-sites $araI_1$ and $araO_2$, respectively, which are separated by 210 base pairs. The N-terminal arm binds to the DNA-binding domain, this forms a covalent connection that hold the DNA-binding domains in a relatively rigid position that favors their binding to the two well separated half-sites. This binding generates a DNA loop, which interferes with the access of the RNA polymerase to the promoter region (**Figure** 4 (b)) (Schleif, 2003; Schleif 2000; Ross *et al.* 2003). When arabinose is present in the cell, it binds to the dimerization pockets in the AraC dimer and its N-terminal arms bind over the arabinose; in turn freeing the DNA-binding domain. Since the dimer is no longer held in the same orientation, the now free DNA-binding domain easily and preferentially binds to the half-site adjacent to $araI_1$ – $araI_2$. In this position the AraC stimulates the transcription from the p_{BAD} promoter (**Figure** 4 (c)) (Schleif, 2003; Schleif 2000; Rodgers and Schleif, 2012).

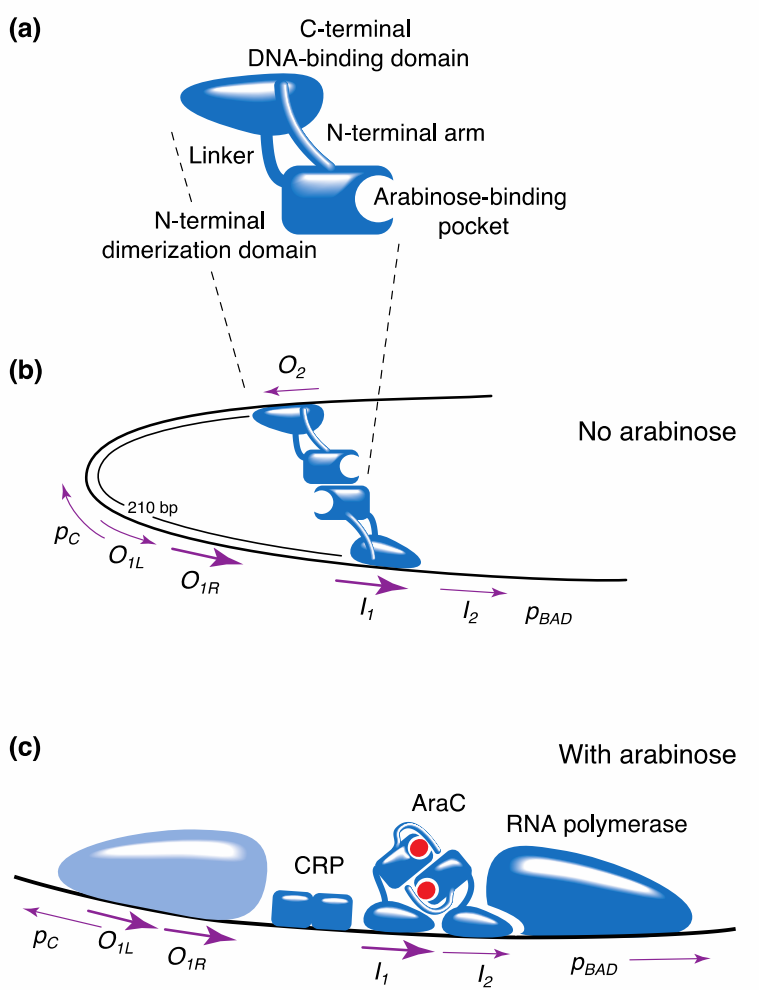


Figure 4 Proposed model for the regulation of the L-arabinose operon by arabinose. (Schleif, 2000)

Previous studies have elucidated the exact binding sites of the AraC dimer to the *araI* operator site. Hendrickson and Schleif showed that the AraC binding sequence has three consensus regions and that these correspond to three adjacent major groove regions of the *araI* operator site (**Figure** 5) (Hendrickson and Schleif, 1985).

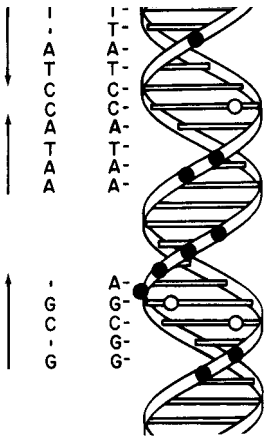


Figure 5 AraC binding to three adjacent major grooves of DNA. White circles represent Guanine contacts; and black circles represent phosphate contacts. The binding sites on *araI* are shown to the left of the DNA helix; and the consensus sequence of the six AraC binding sites is also shown (Hendrickson and Schleif, 1985).

The *araI* operator binding site sequences where then shown to be two direct repeats, the sequences of $araI_1$ and $araI_2$ (**Figure** 6), and via contact probing technique it was then found that last bases in the second direct repeat ($araI_2$) did not have a strong binding contact with AraC (**Figure** 7) (Brunelle and Schleif, 1989).

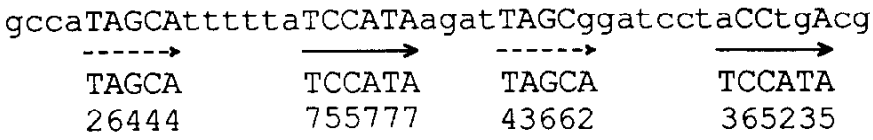


Figure 6 Sequences of the two direct repeats of *araI*. Direct repeat elements are indicated by dashed (first direct repeat) and continuous (second direct repeat) arrows. Capitalized base pairs under the arrows represent the consensus sequences for AraC binding sites (Brunelle and Schleif, 1989).

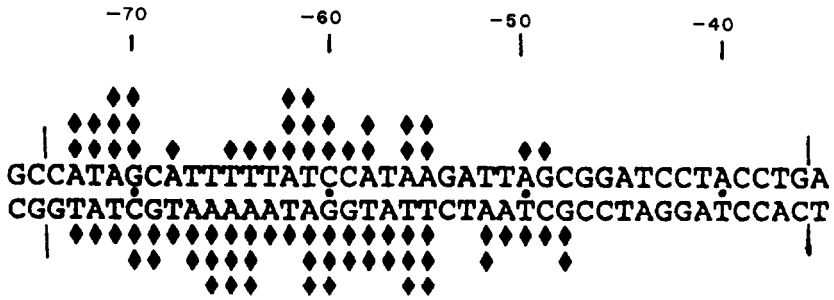


Figure 7 Missing contact data for AraC protein and *araI* operator. The top sequence is wild type AraC protein and the bottom sequence is *araI*. Diamonds represent the intensity of contact between the two regions; one diamond indicates relatively weak effect, and two or three indicate progressively stronger effect (Brunelle and Schleif, 1989).

5. Objectives

Bacteria exert a lot of effort to regulate transcription initiation, not only for economical reasons, as unregulated transcription of genes can use many resources, but also because it is the result of millions of years of evolution driven by survival (Browning and Bubsy, 2004). *Cis*-regulatory elements and transcription factors control most biological processes, however, little is known about the functional and evolutionary effects of epistasis across these molecules (Anderson *et al.*, 2015).

The goal of this work was to elucidate some of these biological questions, by studying the epistatic interactions of the CRE *araI* of the arabinose operon in *E. coli* using a synthetic system. This was achieved by: constructing a plasmid system with the required components; constructing a library of single and double mutants; and testing the epistatic effects in two different environments.

II. Materials and Methods

1. Bacterial Strains, Media and Growth Conditions

The bacterial strains used in this study were electrocompetent Top10F$^-$ *E. coli* and BW25113 cells (CGSC# 7636) (Datsenko and Wanner, 2000). BW25113 are *E. coli* cells that do not have the *araBAD* operon, but have the genes responsible for the transport of extracellular arabinose into the cells – the operon *araFGH* and *araE*. Both strains were made electrocompetent in the lab, flash frozen in liquid nitrogen and kept at -80°C until use (protocol in Appendix 1).

The media used were: Luria-Bertani (LB) (Sigma-Aldrich), which was always supplemented with 50 µg/mL Kanamycin (Kan) antibiotic (Sigma-Aldrich) LB; M9 (1x M9 salts, 0.1mM $CaCl_2$, 2mM $MgSO_4$, 0.2% glucose, Thiamine 1 µg/mL, 2% TritonX and deionized H_2O (dH_2O)) supplemented with 0.01% casamino acids (CAA), 50 µg/mL Kan and 0.2% glycerol; and SOC (Super Optimal Broth (SOB) with glucose) medium (Sigma-Aldrich). When required the media was also supplemented with 0.1% L-arabinose (Sigma).

Unless otherwise stated bacteria were always incubated at 37°C, and if incubated in a shaker incubator, they were shaken at 200 rpm.

2. Plasmid Construction

The plasmid – pBSK-PR-Venus – containing the *E. coli araBAD* operon, the p_{BAD} promoter, and the operator sites (*araI*$_1$, *araI*$_2$, *araO*$_{1L}$, *araO*$_{1R}$, and *araO*$_2$) and the *araC* gene, was used as template to clone the region formerly mentioned. The *araBAD* genes were substituted by the fluorescent protein marker *venus-YFP*. The regions comprising the *venus-YFP* gene (Nagai *et al.*, 2002), the p_{BAD} promoter, the AraC operator sites and the *araC* gene followed by an *E. coli* alpha operon terminator (tL17) were cloned into a low copy number plasmid pZS*, which also contained a Kan resistance gene marker (Lutz and Bujard, 1997). The resulting plasmid, after sequence verification, was used as the wild type for the subsequent experiments.

3. Construction of Mutant Library

A library of single and double mutants in the *araI*$_1$ and *araI*$_2$ regions was constructed as follows.

3.1 Site-directed Mutagenesis

Site-directed mutagenesis was performed using the Quick-Change II™ site-directed mutagenesis kit (Agilent Technologies). The reagents used in the polymerase chain reaction (PCR) are in table 1 and the PCR profile was: 30 seconds (s) at 95°C followed by 12 cycles of 30 s at 95°C, 1 minute (min) at 72°C and 4 min at 68°C, and then 4°C until they were removed from the thermocycler. The PCR products were then cleaned with the UltraClean® GelSpin® DNA Extraction Kit (MO BIO Laboratories, Inc.), digested with Dpn1 restriction enzyme (for 1 hour at 37°C) and then cleaned again with the same kit.

Table 1 Reagents and volumes used in the PCR reactions.

Reagents	**Volume (µL)**
Phusion Polymerase	0.5
Primer Forward	1
Primer Reverse	1
10mM dNTPS	1
DNA template	1
5x Buffer	10
ddH_2O	35.5
Total	50

3.2 Transformation of cells

The mutant plasmids were then transformed into Top10F$^-$ cells via electroporation: 20 µL of the mutant plasmids was mixed with 40 µL of the Top10F$^-$ cells on ice; this mixture was then transferred to an ice-cold electroporation cuvette and the cells were shocked by a pulse of 1800 V. Quickly after the pulse 1mL of warm (≈37°C) SOC medium was added to the cells, they were pipetted into a 1mL eppendorf tube and incubated in a table-top incubator for 1 hour at 37°C and 600 rpm to recover.

3.3 Selection of transformants

After the incubation period, cells were centrifuged at 14000 g for 1 min, most of the supernatant (SN) was discarded and the cells were resuspended in the remaining SN. Cells were plated on a LB agar plate containing Kan, and incubated overnight (≈ 16h). The resulting colonies were picked and streaked onto an LB agar plate supplemented Kan and 0.1% of arabinose and grown overnight. Next, the plates were photographed under fluorescent light and the fluorescent colonies were signaled.

3.4 Sequencing of mutants

One fluorescent colony from each mutant was used to inoculate 1mL of LB with Kan and left incubating overnight in a shaking incubator. Following that plasmids were extracted from cells using the Zippy™ Plasmid Miniprep Kit (Zymo Research), and 10 µL of the product was sent to LGC Genomics (Germany) for Sanger sequencing.

3.5 Maintenance of mutant library

After sequence verification, the plasmids with the desired mutations were transformed into electrocompetent BW25113 cells, as before. BW25113 cells containing the mutagenized plasmids were grown on 1% LB agar plates with Kan overnight. One colony of each mutant was then grown overnight in LB Kan, and the cells were concentrated by centrifugation at 14000 g for 1 min, most of the SN was discarded. In a cryo tube 700 µL of the cells was mixed with 300 µL of 50% glycerol solution and frozen at -80°C.

4. Fitness Assays

From the frozen stock, a sample of each mutant was streaked on an LB agar with Kan plate and incubated overnight.
A colony of each mutant was grown overnight in LB medium with Kan in a shaker incubator. Next 1 µL of each culture was used to inoculate 2 mL of M9 medium with either 0.1% arabinose or no arabinose; four replicates of each mutant for the two different conditions were generated. These were then incubated overnight at 37°C.

After the incubation period, the cultures were submitted to a serial dilution and a sample of each was added to 1.2 mL of the corresponding M9 media (with or without arabinose), and grown for four hours in the same conditions as mentioned above. Regular absorbance measurements were taken, and when the cultures had an OD_{600} of 0.05, 150 µL aliquots were transferred to a black 96 well plate with clear bottom. The fluorescence (500 and 529nm, emission and excitation respectively) and absorbance (OD_{600}) of the cultures were measured in a plate reader (Biotek H1).

5. Statistical Analyses

5.1 Expression Analysis

The fluorescence for each mutant was measured and the results were normalized by the fluorescence of the wild type, so that the fluorescence recorded was relative to the wild type strain in the respective environments. Error propagation was taken into account when calculating the standard deviation of the mean fluorescence for each mutant, because of the intrinsic variation of each replicate when they are normalized by the wild type.

Analysis of variance was calculated to see the effect of each different mutant on expression using the ANOVA function of StatPlus:mac; in this case the response variable was the relative fluorescence, the mutants were the fixed factor, and the different replicates the random factor. Comparisons between the mean fluorescence of each mutant to that of the wild type were done by false discovery rate (FDR) corrected *t*-tests.

5.2 Estimation of Epistasis

Epistasis was calculated using a multiplicative analysis model. This model assumes that the different mutations are not independent; and the formula was: $\varepsilon=\omega_{D1,2} - (\omega_{S1} \times \omega_{S2})$, where $\omega_{D1,2}$ is the relative fluorescence of the double mutants, and ω_{S1} and ω_{S2} are the relative fluorescence of its corresponding single mutants (Cordell, 2002). FDR-correct *t*-tests were done to test if the estimated epistasis were statistically significant.

To assess the effect of the different environments (presence or absence of arabinose) on epistasis – genotype-by-genotype-by-environment (GxGxE) – the analysis of variance was tested, using the replicates as a random factor. Using FDR-

corrected *t*-tests, with the mean epistasis as a response variable, the difference in the magnitude and sign of epistasis was compared between double mutants depending on the location of its mutations.

III. Results

1. Assembly of mutant library

A plasmid containing the regulatory regions of the *araBAD* operon, but with a fluorescent report gene, *venus-yfp*, instead of *the araA, araB* and *araD* genes was constructed for this study and used as template for site-directed mutagenesis to create the mutant library (**Figure** 8).

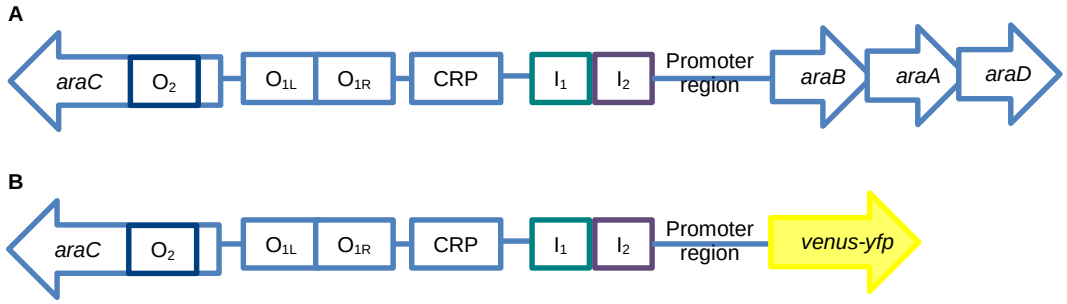


Figure 8 Architecture of the *araBAD* operon and its operators and of the synthetic construct . A. Structure of the operon region used as template for synthetic construct. B. Structure of the wild type plasmid contruct.

Single and double point mutations were introduced in the $araI_2$ and $araI_1$ half-sites of the AraC operator. In total there were 17 single mutants produced, 10 mutants with single mutations in $araI_1$ and 7 mutants with single mutations in $araI_2$ (Table 1). The location of the mutations was based on previous studies that revealed the exact locations the AraC protein dimer binds to the DNA; the bases introduced were chosen at random but keeping the number of transition and transversions at a 1:2 ratio. Then these single mutants were used as templates to create the 20 double mutants; 5 of these had both mutations in $araI_1$, 5 with both mutations in $araI_2$, and 10 mutants with one mutation in $araI_1$ and one mutation in $araI_2$ (Table 2); the combinations of double mutants were chosen randomly.

Table 2 Single Mutant library – Areas highlighted in teal and purple are the $araI_1$ and $araI_2$ regions, respectively.

Mutant	**Sequence**
Single Mutant 1	CGGCGTCACACTTTGCATAGCCAGAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 2	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 3	CGGCGTCACACTTTGCATAGCCATAGGATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 4	CGGCGTCACACTTTGCATAGCCATAGCCTTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 5	CGGCGTCACACTTTGCATAGCCATAGCATTTTTAGCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 6	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATTCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 7	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCAATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 8	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCGTAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 9	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCAAAAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 10	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATTAGATTAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 11	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATGAGCGGATCCTACCTGACGCTTTTTA
Single Mutant 12	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTCGCGGATCCTACCTGACGCTTTTTA
Single Mutant 13	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTACCGGATCCTACCTGACGCTTTTTA
Single Mutant 14	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGTGGATCCTACCTGACGCTTTTTA
Single Mutant 15	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAGCTGACGCTTTTTA
Single Mutant 16	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACATGACGCTTTTTA
Single Mutant 17	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGTCGCTTTTTA
Wild Type	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA

Table 3 Double Mutant library – Double mutants are named according to the two single mutants it's composed of, i.e. D1,6 means that it is made of single mutants 1 and 6, respectively. Areas highlighted in teal and purple are the $araI_2$ and $araI_1$ regions, respectively.

Mutant	Sequence
Double Mutant – D1,6	CGGCGTCACACTTTGCATAGCCAGAGCATTTTTATTCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Double Mutant – D1,2	CGGCGTCACACTTTGCATAGCCAGACCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Double Mutant – D5,7	CGGCGTCACACTTTGCATAGCCATAGCATTTTTAGCAATAAGATTAGCGGATCCTACCTGACGCTTTTTA
Double Mutant – D3,9	CGGCGTCACACTTTGCATAGCCATAGGATTTTTATCCAAAAGATTAGCGGATCCTACCTGACGCTTTTTA
Double Mutant – D4,10	CGGCGTCACACTTTGCATAGCCATAGCCTTTTTATCCATTAGATTAGCGGATCCTACCTGACGCTTTTTA
Double Mutant – D12,14	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTCGTGGATCCTACCTGACGCTTTTTA
Double Mutant – D11,13	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATGACCGGATCCTACCTGACGCTTTTTA
Double Mutant – D14,17	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGTGGATCCTACCTGTCGCTTTTTA
Double Mutant – D16,17	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACATGTCGCTTTTTA
Double Mutant – D13,17	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTACCGGATCCTACCTGTCGCTTTTTA
Double Mutant – D1,13	CGGCGTCACACTTTGCATAGCCATGGCATTTTTATCCATAAGATTACCGGATCCTACCTGACGCTTTTTA
Double Mutant – D2,15	CGGCGTCACACTTTGCATAGCCATACCATTTTTATCCATAAGATTAGCGGATCCTAGCTGACGCTTTTTA
Double Mutant – D2,16	CGGCGTCACACTTTGCATAGCCATACCATTTTTATCCATAAGATTAGCGGATCCTACATGACGCTTTTTA
Double Mutant – D5,13	CGGCGTCACACTTTGCATAGCCATAGCATTTTTAGCCATAAGATTACCGGATCCTACCTGACGCTTTTTA
Double Mutant – D5,15	CGGCGTCACACTTTGCATAGCCATAGCATTTTTAGCCATAAGATTAGCGGATCCTAGCTGACGCTTTTTA
Double Mutant – D7,14	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCAATAAGATTAGTGGATCCTACCTGACGCTTTTTA
Double Mutant – D8,16	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCGTAAGATTAGCGGATCCTACATGACGCTTTTTA
Double Mutant – D8,17	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCGTAAGATTAGCGGATCCTACCTGTCGCTTTTTA
Double Mutant – D9,15	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCAAAAGATTAGCGGATCCTAGCTGACGCTTTTTA
Double Mutant – D9,17	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCAAAAGATTAGCGGATCCTACCTGACGCTTTTTA
Wild Type	CGGCGTCACACTTTGCATAGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTA

2. Characterization of mutational effects on fitness

To test the effect of the mutations on fitness, in this case measured as relative fluorescence, wild type and related mutants were grown in media containing the operon inducer arabinose, or in media without it (**Figures** 9 and 10). The results indicate that mutations in $araI_1$ and $araI_2$ operator sites significantly alter the expression of *venus-yfp*, in both the presence and absence of arabinose in the growth media.

In the presence of arabinose, the AraC protein binds to both $araI_1$ and $araI_2$ operators half-sites; therefore mutations in these sites were expected to affect gene expression. Mostly, as expected, relative fluorescent levels were significantly lower than that of the wild type, except for three single mutants. The relative fluorescent levels of $araI_1$ single and double mutants, excluding single mutants 1 and 4, which had a higher expression, decreased significantly for all but the single mutant 10. The same pattern was seen for single and double mutants of $araI_2$, excluding single mutant 17, which surprisingly also had higher levels of relative fluorescence. Double mutants with mutations in the $araI_1$ and $araI_2$ half-sites, had overall the lowest levels of expression, with exception of D2,16; the levels of expression were also more similar between these mutants. These results illustrate that mutations in certain base pairs affect the binding of the AraC dimer to the DNA more than others; this can especially be seen when the expression of the double mutants is not necessarily lower or either a combination of the levels of expression of its constituent single mutants.

In the absence of arabinose, AraC protein binds to $araI_1$ and $araO_2$. Here mutations in the $araI_2$ half-site should not alter expression, and overall it was found that it didn't; only two single mutants' expression was significantly increased from that of the wild type, and also only two double mutants with both mutations in the $araI_2$ half-site were significantly different (increased expression) from the wild type. Relative fluorescence levels did not vary greatly between single mutants or between double mutants; and also the relative fluorescence levels did not differ much in regards to the number of mutants. Mutations in the $araI_1$, in this environment (no arabinose), should, however, disturb the expression, as AraC binds strongly to $araI_1$ (Schleif and Brunelle, 1989). The results concur with these expectations; in this case of the 10 single mutants, 6 showed a significant increase of relative fluorescence levels, and all of the double mutants with both mutations in $araI_1$ were also significantly higher than the wild type. The expression levels of double mutants with mutations in $araI_1$ and $araI_2$ should look

more similar to those with both mutations in *araI*$_1$, since *araI*$_2$ does not play a regulatory part here, however the relative fluorescence was diverse for the different mutants, only half showed a significant increase in relative fluorescence, indicating that perhaps the mutations in *araI*$_2$ could be affecting binding of the AraC dimer to this region.

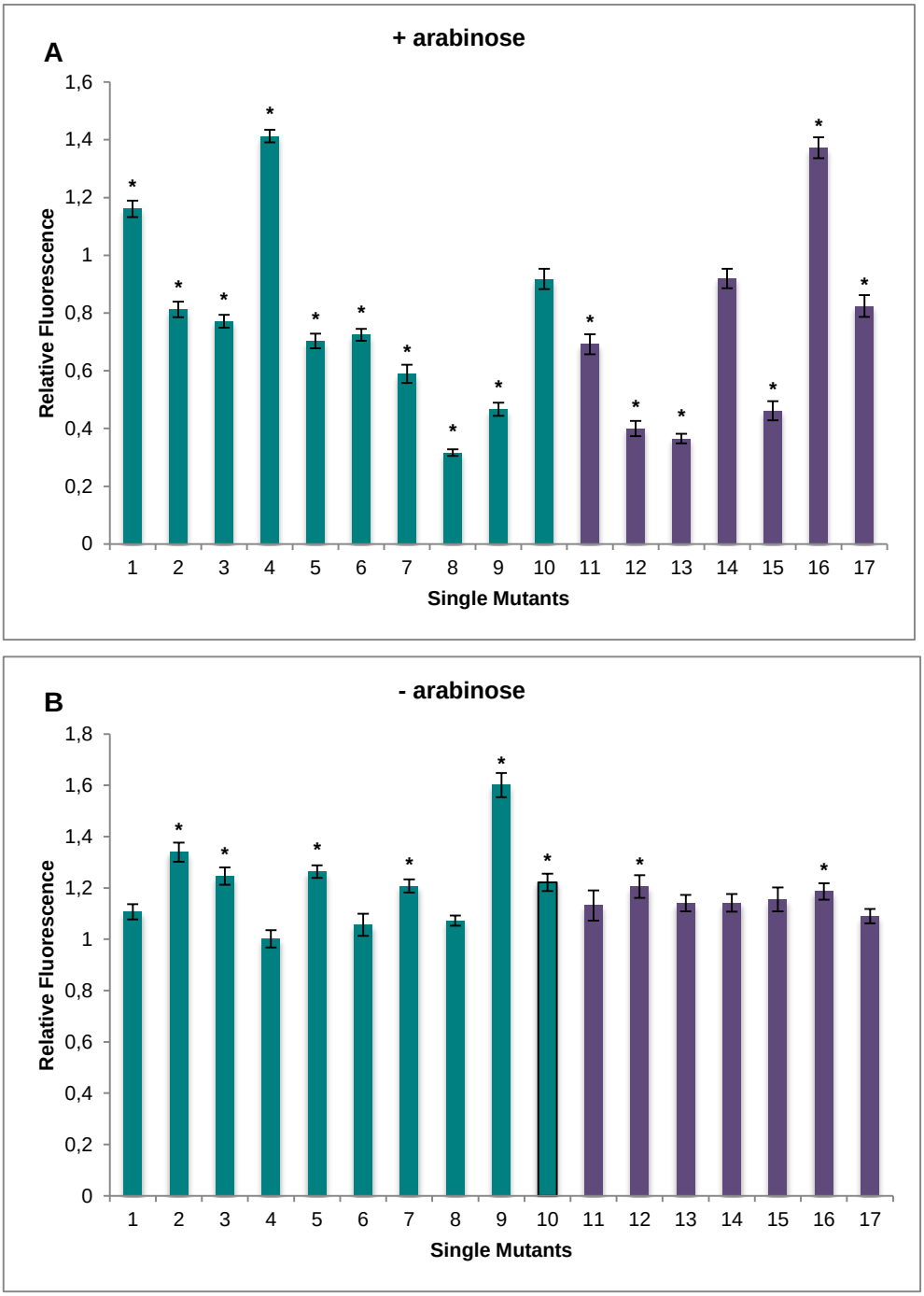


Figure 9 Relative fluorescence of single mutants in the presence and absence of arabinose (**A** and **B** respectively). Bars represent the fluorescence relative to the wild type. Teal colored bars represent mutants in the *araI*$_1$ operator; purple bars represent mutants in the *araI*$_2$ operator. Fluorescence was normalized to the wild type, which is 1. Error bars are standard errors of the mean. Stars denote mutants that significantly differ from the wild type.

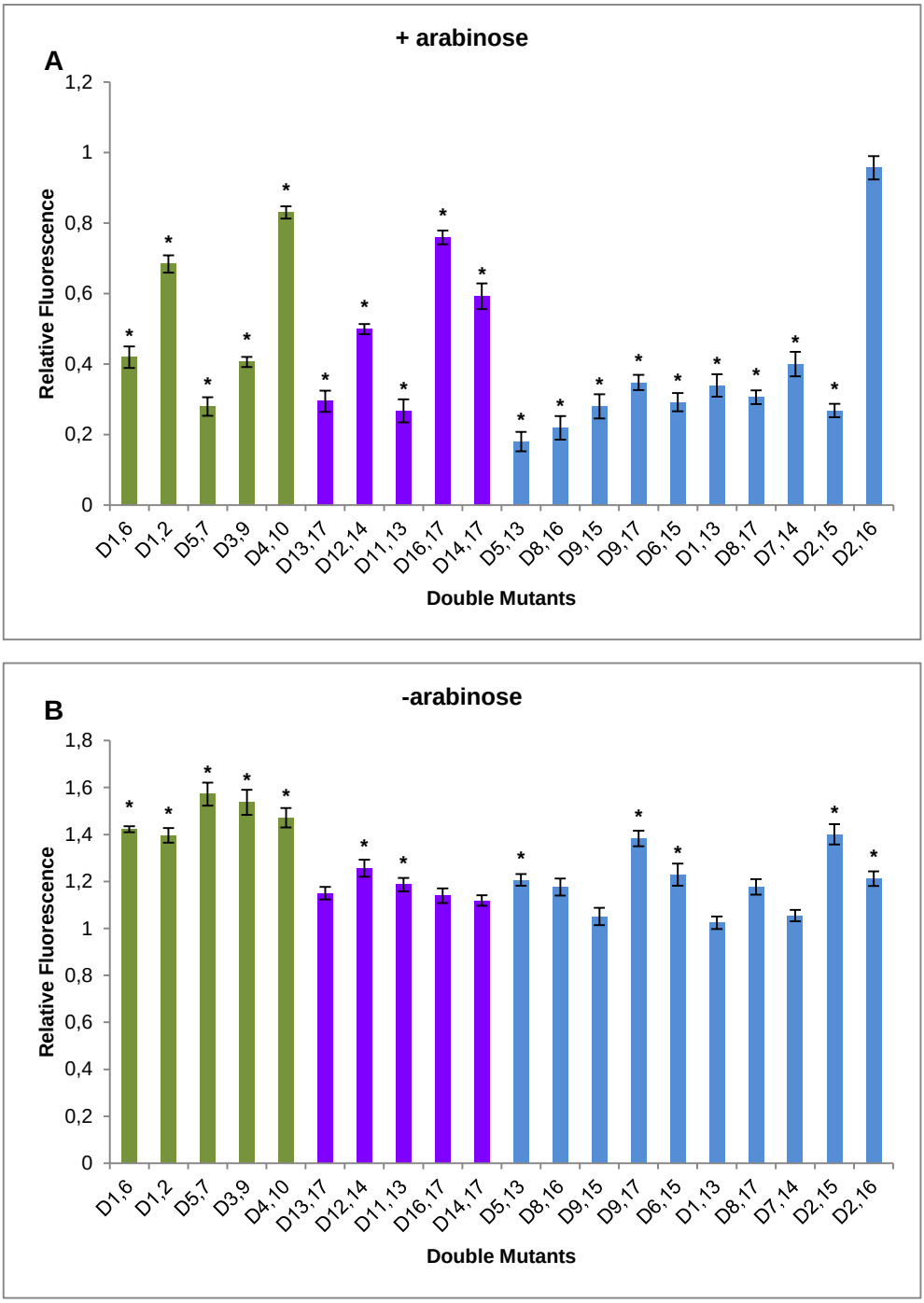


Figure 10 Relative fluorescence of double mutants in the presence and absence of arabinose (**A** and **B** respectively). Bars represent the fluorescence relative to the wild type. Green bars represent mutants with both mutations in the *araI*$_1$ operator; violet bars represent mutants with both mutations in the *araI*$_2$ operator; blue bars represent mutants with mutations in *araI*$_1$ and *araI*$_2$ operators. Fluorescence was normalized to the wild type, which is 1. Error bars are standard errors of the mean. Stars denote mutants that significantly differ from the wild type.

3. Characterization of epistatic interactions

3.1 Epistatic interactions in different environments

To evaluate effect of the environment on epistasis, the epistatic interactions were measured as the deviation from the multiplicative expectation of gene expression levels based on the single mutant effects. The results illustrate that there are significant epistatic interactions in both the presence and absence of arabinose (**Figure** 11); and that these are also environment dependent was shown by the high GxGxE interaction ($F_{19,120}$=21.51, $p < 0.0001$) .

In the presence of arabinose, of the 10 double mutants that show significant epistasis, only one showed synergistic epistasis (where the combined effect of mutations is greater together than the sum of their individual effects), the remainder interactions were of antagonistic epistasis (where the combined effect is less than the sum of the effect of individual mutations).

When there is no arabinose in the media, most epistatic interactions were of a negative nature (8 of 10 double mutants), and only two double mutants demonstrating positive epistasis.

Sign epistasis was also observed in two double mutants (D1,6 and D4,10, or 20 and 24 in **figure 11**) when exposed to a different environment, negative in the presence of arabinose and positive in its absence (**Figure** 11).

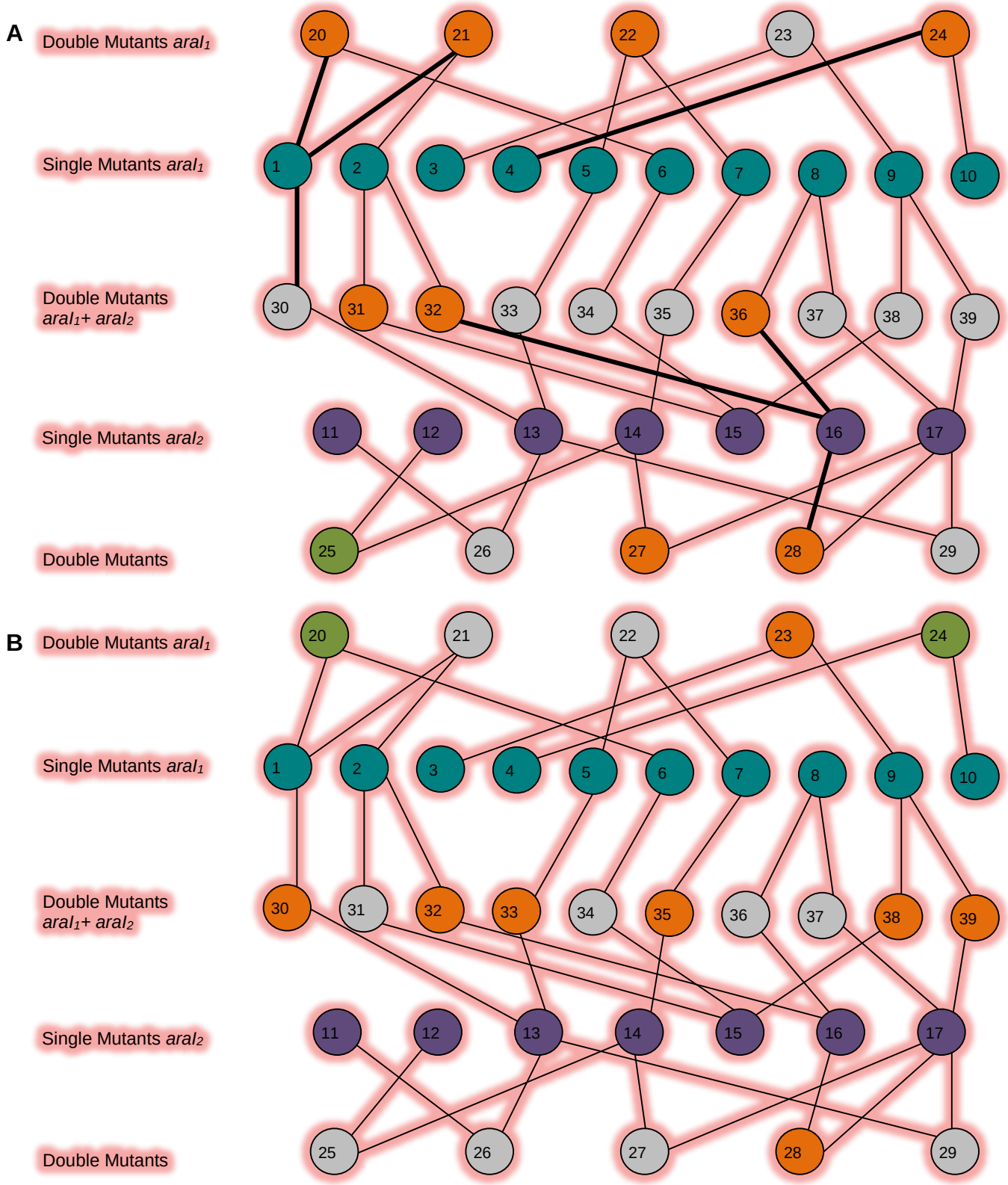


Figure 11 Network of mutant epistatic interactions. **A**. Epistatic interaction network in the presence of arabinose. **B**. Epistatic interaction network in the absence of arabinose. Single mutants are in teal and purple circle and are connected by lines to the corresponding double mutants that contain them; numbers correspond to mutants according to Table 2. Double mutants are in: grey (non interacting mutations), orange (significant negative epistasis) and green (significant positive epistasis) circles; numbers inside correspond to double mutants as follows: 20 (D1,6), 21 (D1,2), 22 (D5,7), 23 (D3,9), 24 (D4,10), 25 (D12,14), 26 (D11,13), 27 (D14,17), 28 (D16,17), 29 (D13,17), 30 (D1,13), 31 (D2,15), 32 (D2,16), 33 (D5,13), 34 (D6,15), 35 (D7,14), 36 (D8,16), 37 (D8,17), 38 (D9,15), and 39 (D9,17). Thick connecting lines indicate sign epistasis.

3.2 Epistatic interactions relative to physical location of mutations

To see if the epistatic interactions of the double mutants were dependent on the physical location of their mutations on the half-site operators, the analysis of variance was calculated. The results show that there is significant difference between the operators and that this is also dependent on the presence ($F_{2,17}$=25.083, $p < 0.0001$) and absence ($F_{2,17}$=39.089, $p < 0.0001$) of arabinose.

In the presence of arabinose, mutations solely in the *araI₁* resulted in a much higher negative epistatic effect compared to when both mutations were on the *araI₂* operator (t_8=-3.257, $p < 0.05$) or when there was one mutation in both *araI₁* and *araI₂* operator half-sites (t_{13}=-4.304, $p < 0.001$) (**Figure** 12 **A**).

In the absence of arabinose in the media, the greatest negative epistatic interactions were found when double mutants had one mutation in each operator, whereas double mutants with both mutations in *araI₁* had a much lower epistatic effect (t_{13}=4.366, $p < 0.001$) in comparison, as well as double mutants with both mutations in *araI₂* (t_{13}=2.165, $p < 0.05$) (**Figure** 12 **B**).

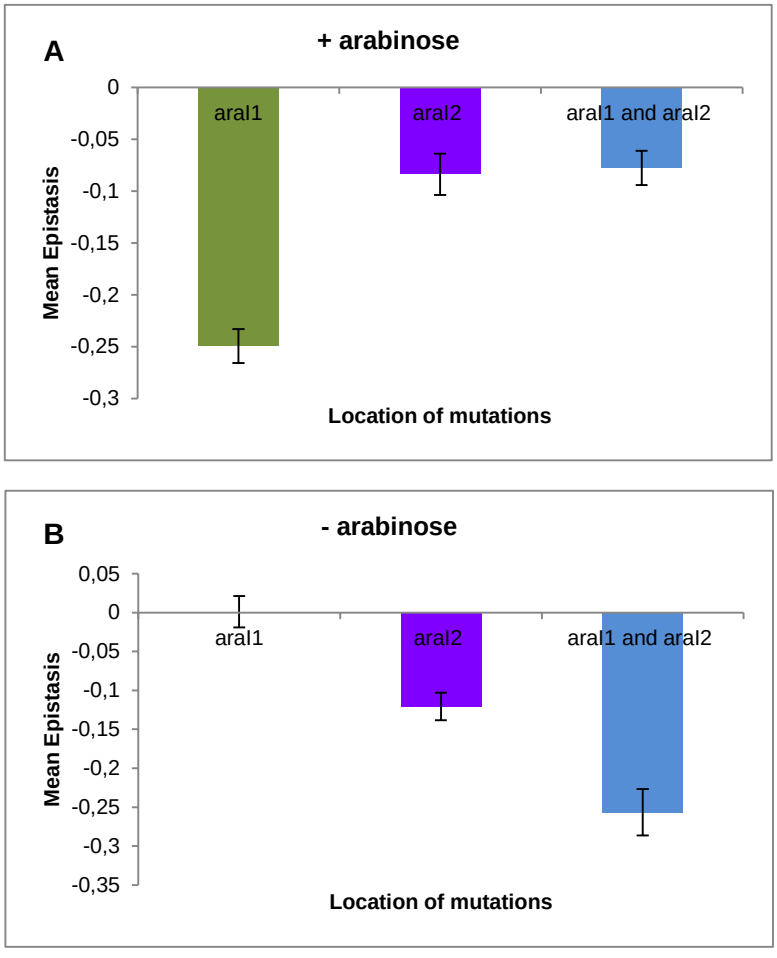


Figure 12 Mean epistasis of double mutants relative to their location on the operator half-sites in the presence (**A**) and absence (**B**) of arabinose. Green bars represent mutants with both mutations in the *araI₁* operator; violet bars represent mutants with both mutations in the *araI₂* operator; blue bars represent mutants with mutations in *araI₁* and *araI₂* operators. Error bars are standard deviation.

IV. Discussion

In this work a synthetic system was generated and used to study the epistasis of CREs. The system uses genes of the arabinose operon of *E. coli*, specifically the *araI* operator site, made of the two adjacent regions $araI_1$ and $araI_2$, to which AraC protein binds to regulate expression of the p_{BAD} promoter of the *araBAD* operon. To analyze the epistatic interactions in these operator half-sites, a library of 17 single mutants, with 10 mutations in $araI_1$ and 7 mutations in $araI_2$, and a library of 20 double mutants, consisting of 5 double mutants with both mutations in $araI_1$, 5 double mutants with both mutations in $araI_2$, and 10 double mutants with one mutation in each operator half-site, were created. Further the epistatic interactions between the double mutants were tested in two environments: (1) in the presence or (2) in the absence of the operon inducer arabinose.
The results demonstrate that there were significant epistatic interactions in between the mutants of these operator sites, and that these are highly dependent on the environment in which they are present.

Using a synthetic system to study epistasis is a relatively simple and suitable method, because it allows a simple estimation of how the mutations affect gene expression, and therefore epistasis, without it being affected by other cellular processes (Sanjuán *et al*., 2004). In this case the *araBAD* were replaced for a fluorescent reporter gene – *venus*-yfp; and so the gene expression of the mutants was measured relative fluorescence levels. The drawback of this, however, is that the amount of gene product depends on the plasmid number. In this case the low copy number plasmid pZs* was used; this will generate a certain amount of transcripts that is susceptible to variation. To further quantify the plasmid number for all the mutants qPCR should be used, this would make the results more comparable (Chou and Marx, 2012). If the construct were in the chromosome, an even more precise estimate would have been done; allowing comparisons with other studies. The difference in experimental systems and mutations involved has been one of the main problems in being able to make general conclusions on studies of epistasis (Schenk *et al.*, 2013; Sanjuán et al., 2004).

The single mutations were designed based on previous studies that showed where the AraC protein dimer DNA binding domains' bind in that DNA sequence (Brunelle

and Schleif 1989; Hendrickson and Schleif, 1985). These studies also show that there is a consensus-binding region in these AraC DNA binding domains. Further studies could include all the possible different bases at each point mutation, and analyze how changing from a purine to a pyrimidine affects the binding and gene expression. Possibly changing from one purine to another would not alter gene expression as much, as it is known that they have more similar structures.

In Niland *et al.* (1996), it was found that single point mutations in $araI_1$, reduced AraC binding by tenfold in the presence of arabinose and a sixfold decrease in its absence, whilst substitutions at other bases had little to no effect. In this study it was shown that mutations in the $araI_1$, overall significantly decrease gene expression in the presence of arabinose. These results corroborate with the position weight matrix for AraC binding from the RegulonBD (Salgado *et al.* 2013) and with the fact that this site is highly conserved in the *Enterobacteriaceae* family. However in the absence of arabinose, gene expression levels increased for both single and double mutants in the $araI_1$ region. Although these results contrast with other published studies, it should be taken into account that in the study of Niland *et al.* (1996) the binding of AraC to the operator was experimentally inferred by gel mobility shift assays. Such method does not consider *in vivo* processes, such as DNA looping caused by the AraC dimer and genetic noise that may influence binding. This increase in expression levels could be explained by the fact that in the absence of arabinose the promoter is being repressed, the AraC dimers bind to $araI_1$ and $araO_2$, forming a slightly rigid structure that holds the DNA loop and interferes with promoter region where the RNA polymerase would bind to initiate transcription. Since the binding of AraC to $araI_1$ is reduced, this could create a shift in the way the AraC dimer is held in place or change its binding affinity to the $araI_2$ operator site instead (since it is less conserved), thus allowing the RNA polymerase to bind to the p_{BAD} promoter (**Figure** 13).

Mutations in the $araI_2$ region did not influence gene expression in the absence of arabinose as it is expected, since the AraC dimer does not bind in that region in the absence of arabinose. When arabinose was present in the media, gene expression decreased greatly, possibly because the AraC dimer preferentially binds to $araO_2$ instead. This could explain the variability of gene expression between the mutants and why some mutations affect gene expression less than others, once the $araI_2$ direct repeats are much less conserved. Another explanation is that these mutations induced a conformational change in the DNA structure that affects the binding of the AraC dimer to the $araI_1$ region (**Figure** 13).

Double mutants with one mutation in each operator half-site, significantly decreased gene expression in all but one mutant in the presence of arabinose; although these results should look more similar to those of single or double mutants in $araI_1$, since $araI_2$ does not play a role in AraC binding in this environment, the relative fluorescence levels are some the lowest, this supports the idea that mutations in $araI_2$ alter the binding of the AraC to $araI_1$.

The effects of the environment in epistasis have been widely described in the literature. Most of these studies, however, focus on proteins and environments that cause a complex cellular response (Wang *et al.*, 2012; You and Yin, 2012; de Vos *et al.*[1]; 2013). The different environments used in this study – the presence or absence of arabinose in the media – only caused a shift in binding affinity of the AraC dimer's DNA binding domain from one operator site to another, due to the binding of arabinose to the arabinose-binding pocket in the AraC dimer. Despite being a rather simple and controllable change in environment, it proved enough to alter the epistatic interactions between the different mutants (**Figures** 11 and 12).

The results also demonstrate that epistatic interactions between the double mutants were dependent on their physical location, i.e. in which operator site(s) the mutations were in (**Figure** 12). In the presence of arabinose, when the AraC-arabinose complex binds to both *araI* operator half-sites, mutations in $araI_1$ showed greater negative epistasis than mutations in $araI_2$; this is in accordance with previous studies that showed a greater binding affinity to $araI_1$, and that $araI_1$ is more conserved, so mutations in this region lead to a greater negative response (Brunelle and Schleif, 1989, Niland *et al.*, 1996). In the absence of arabinose, the AraC dimer does not bind to $araI_2$, so the greater negative epistatic effect when both mutations are in this operator is surprising; as is the even higher negative epistatic effect when there is one mutation in each operator half-site. These results lead to conclude that mutations in $araI_2$ do interfere with AraC binding to $araI_1$; this has been demonstrated in studies where there are mutations in the flanking regions of the transcription factor binding sites (Levo and Segal, 2014). To see if these interferences were simply caused by physical proximity of the two half-sites, the same experiment could be done with mutations in the $araO_2$ operator site instead, since it is more distant to the $araI_1$ region, and see how this affected AraC binding.

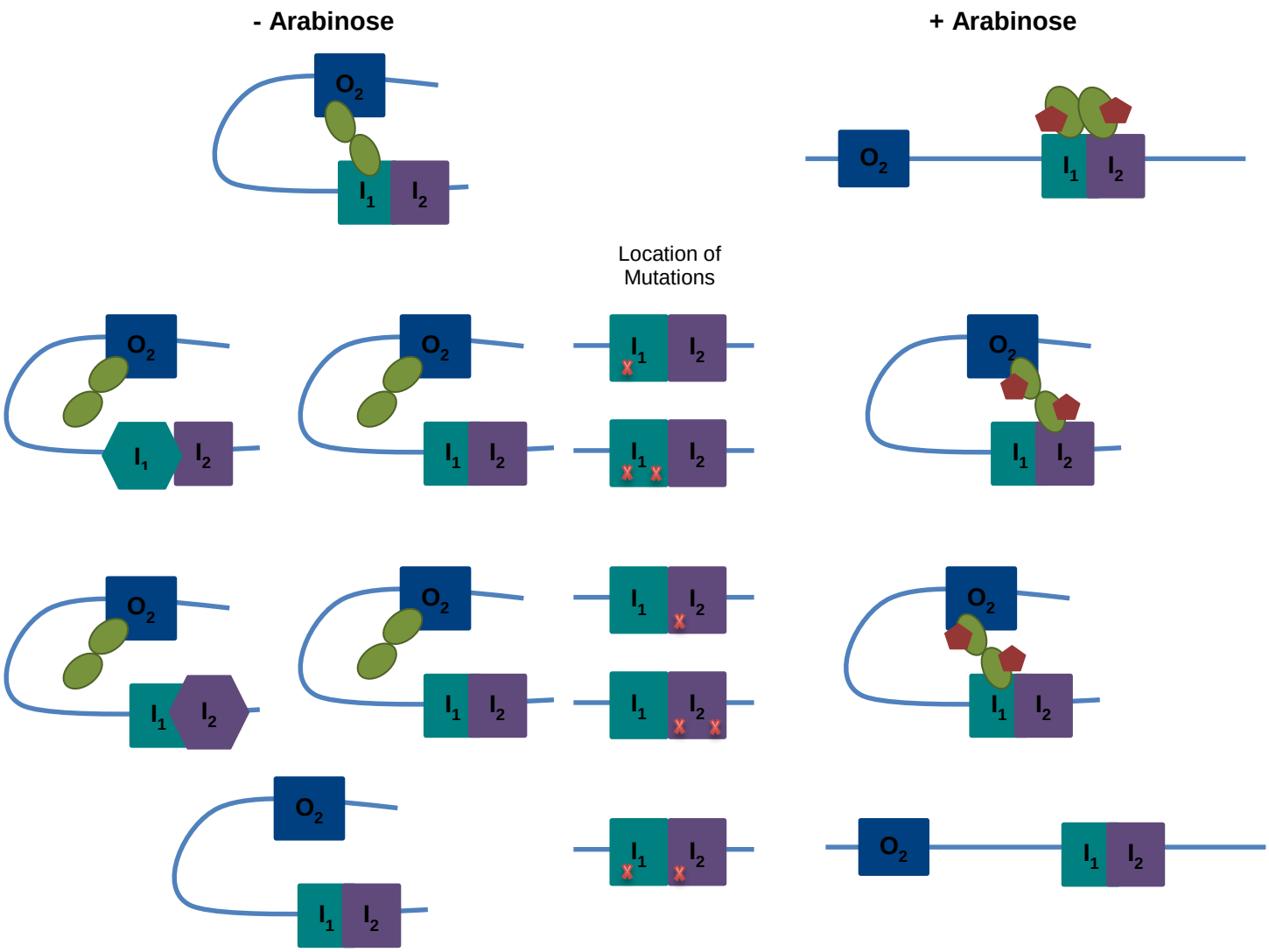
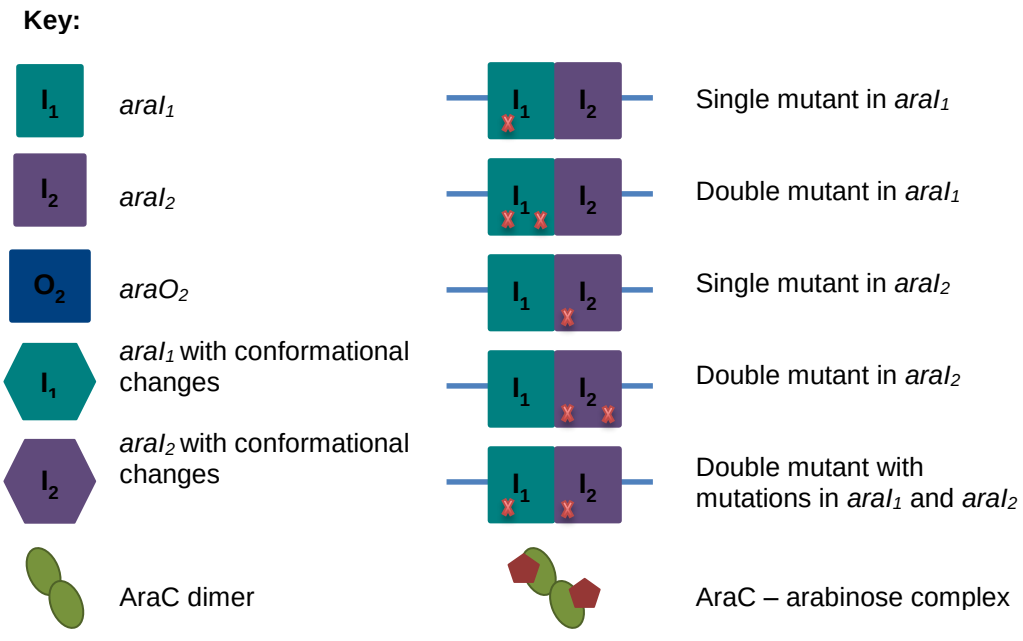


Figure 13 Possible binding outcomes caused my mutations in *araI₁* and *araI₂*.



The importance of studying CREs and epistasis among these has been recognized (Jolma *et al*. 2013). Not only are these important to understand evolutionary history of regulatory elements, they could provide insight of fitness landscapes; and also lead to understanding all the properties of a network, which could result in optimal design of synthetic constructs.

The results reported herein suggest that: (1) greater alterations in gene expression happen when the mutations in the CRE are in a region that is highly conserved and has greater binding affinity to the transcription factor; whilst (2) less variation in the gene expression is found when there are mutations in different operator sites. It was also demonstrated here that epistatic interactions are highly dependent on the environment they are in.

References

Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M.B., 2010. Annotating non-coding regions of the genome. *Nature reviews. Genetics*, 11(8), pp.559–571.

Azevedo, R.B.R., Lohaus, R., Srinivasan, S., Dang, K.K. & Burch, C.L., 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature*, 440(7080), pp.87–90.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.-F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R. & Bulyk, M.L., 2009. Diversity and complexity in DNA recognition by transcription factors. *Science*, 324(5935), pp.1720–3.

Bateson, W. (1909). *Mendel's principles of heredity*. Cambridge Univ. Press, Cambridge.

Browning, D.F. & Busby, S.J., 2004. The regulation of bacterial transcription initiation. *Nature reviews. Microbiology*, 2(1), pp.57–65.

Brunelle, A. & Schleif, R., 1989. Determining residue-base interactions between AraC protein and araI DNA. *Journal of Molecular Biology*, 209(4), pp.607–622.

Chou, H.H. & Marx, C.J., 2012. Optimization of Gene Expression through Divergent Mutational Paths. *Cell Reports*, 1(2), pp.133–140.

Datsenko, K. a & Wanner, B.L., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp.6640–6645.

[1]De Vos, M.G.J., Poelwijk, F.J. & Tans, S.J., 2013. Optimality in evolution: New insights from synthetic biology. *Current Opinion in Biotechnology*, 24(4), pp.797–802.

[2]De Vos, M.G.J., Poelwijk, F.J., Battich, N., Ndika, J.D.T. & Tans, S.J., 2013. Environmental Dependence of Genetic Constraint. *PLoS Genetics*, 9(6).

Elena, S.F. & Lenski, R.E., 1997. Test of synergistic interactions among deleterious mutations in bacteria. *Nature*, 390(6658), pp.395–398.

Englesberg, E. & Wilcox, G., 1974. Regulation : Positive Control. *Annual review of genetics*, 8(17), pp.219–242.

Englesberg, E., 1961. Enzymatic Characterization of 17 L-arabinose Negative Mutants of Escherichia coli. *Journal of Bacteriology*, 81(6), pp.996 – 1006.

Englesberg, E., 1971. *Metabolic Regulation*. Academic Press, New York, pp. 257–296**.**

Gielow, L., Largen, M. & Englesberg, E., 1971. Initiator constitutive mutants of the L-arabinose operon (OIBAD) of Escherichia coli B/r. *Genetics*, 69(3), pp.289–302.

Greenblatt, J. & Schleif, R., 1971. Arabinose C protein: regulation of the arabinose operon in vitro. *Nature: New biology*, 233(40), pp.166–170.

Gross, J. & Englesberg, E., 1959. Determination of the order of mutational sites governing L-arabinose utilization in Escherichia coli B/r bv transduction with phage Plbt. *Virology*, 9, pp.314–331.

Hartl, D.L., 2014. What can we learn from fitness landscapes? *Current opinion in microbiology*, 21, pp.51–57.

Hendrickson, W. & Schleif, R., 1985. A dimer of AraC protein contacts three adjacent major groove regions of the araI DNA site. *Proceedings of the National Academy of Sciences of the United States of America*, 82(10), pp.3129–3133.

Hueck, C.J. & Hillen, W., 1995. Catabolite repression in Bacillus subtilis: a global regulatory mechanism for the gram-positive bacteria? *Molecular microbiology*, 15(3), pp.395–401.

Ibáñez-Marcelo, E. & Alarcón, T., 2014. The topology of robustness and evolvability in evolutionary systems with genotype-phenotype map. *Journal of theoretical biology*, 356, pp.144–62.

Jacob, F., Perrin, D., Sanchez, C. and Monod, J. (1960). L'óperon: groupe de genes à expression coordonée par un opérateur. *C.R. Acad. Sci. Paris, 250*, 1727-1729.

Jasnos, L. & Korona, R., 2007. Epistatic buffering of fitness loss in yeast double deletion strains. *Nature genetics*, 39(4), pp.550–554.

Jeziorska, D.M., Jordan, K.W. & Vance, K.W., 2009. A systems biology approach to understanding cis-regulatory module function. *Seminars in Cell and Developmental Biology*, 20(7), pp.856–862.

Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E. & Cooper, T.F., 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332(6034), pp.1193–1196.

Kussell, E., 2013. Evolution in microbes. *Annual review of biophysics*, 42, pp.493–514.

Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C. & Cohen, B.A., 2012. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47), pp.19498–503.

Lalic, J. & Elena, S.F., 2012. Epistasis between mutations is host-dependent for an RNA virus. *Biology Letters*, 9, pp.1 –4.

Lalic, J. & Elena, S.F., 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity*, 109(2), pp.71–77.

Lee, N., Francklyn, C. & Hamilton, E.P., 1987. Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 84(24), pp.8814–8.

Lehner, B., 2011. Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, 27(8), pp.323–331.

Lelli, K.M., Slattery, M. & Mann, R.S., 2012. Disentangling the many layers of eukaryotic transcriptional regulation. *Annual review of genetics*, 46, pp.43–68.

Levo, M. & Segal, E., 2014. In pursuit of design principles of regulatory sequences. *Nature reviews. Genetics*, 15(7), pp.453–68.

Lim, H.N., Lee, Y. & Hussein, R., 2011. Fundamental relationship between operon organization and gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 108(26), pp.10626–10631.

Lutz, R. & Bujard, H., 1997. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 25(6), pp.1203–1210.

Maisnier-Patin, S., Roth, J.R., Fredriksson, A., Nyström, T., Berg, O.G. & Andersson, D.I., 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nature genetics*, 37(12), pp.1376–1379.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. & Miyawaki, A., 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology*, 20(1), pp.87–90.

Niland, P., Hühne, R. & Müller-Hill, B., 1996. How AraC interacts specifically with its target DNAs. *Journal of molecular biology*, 264(4), pp.667–674.

Osbourn, A.E. & Field, B., 2009. Operons. *Cellular and molecular life sciences : CMLS*, 66(23), pp.3755–75.

Parera, M., Fernàndez, G., Clotet, B. & Martínez, M.A., 2007. HIV-1 protease catalytic efficiency effects caused by random single amino acid substitutions. *Molecular Biology and Evolution*, 24(2), pp.382–387.

Perez, J.C. & Groisman, E. A., 2009. Transcription factor function and promoter architecture govern the evolution of bacterial regulons. *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), pp.4319–4324.

Phillips, P.C., 2008. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nature reviews. Genetics*, 9(11), pp.855–867.

Poelwijk, F.J., Heyning, P.D., de Vos, M.G., Kiviet, D.J. & Tans, S.J., 2011. Optimality and evolution of transcriptionally regulated gene expression. *BMC Systems Biology*, 5(1), p.128.

Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M. & Tans, S.J., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, 445(7126), pp.383–386.

Poelwijk, F.J., Tănase-Nicola, S., Kiviet, D.J. & Tans, S.J., 2011. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology*, 272(1), pp.141–144.

Ralston, A. (2008) Operons and prokaryotic gene regulation. Nature Education 1(**1**):216.

Ramos, J.L., García-Salamanca, A., Molina-Santiago, C., & Udaondo, Z., 2013. Operon. *Brenner's Encyclopedia of Genetics*, Second edition, Volume 5, pp.176-180.

Rodgers, M.E. & Schleif, R., 2012. Heterodimers reveal that two arabinose molecules are required for the normal arabinose response of Arac. *Biochemistry*, 51(41), pp.8085–8091.

Ross, J.J., Gryczynski, U. & Schleif, R., 2003. Mutational analysis of residue roles in AraC function. *Journal of Molecular Biology*, 328(1), pp.85–93.

Sa-Nogueira, I., Nogueira, T. V, Soares, S. & Lencastre, H. De, 1997. Nucleotide Sequence , Genetic Organization and Expression. *Microbiology*, 143, pp.957–969.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernández, S., Alquicira-Hernández, K., López-Fuentes, A., Porrón-Sotelo, L., Huerta, A.M., Bonavides-Martínez, C., Balderas-Martínez, Y.I., Pannier, L., Olvera, M., Labastida, A., Jiménez-Jacinto, V., Vega-Alvarado, L., Del Moral-Chávez, V., Hernández-Alvarez, A., Morett, E. & Collado-Vides, J., 2013. RegulonDB v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Research*, 41, pp.203–213.

Salverda, M.L.M., Dellus, E., Gorter, F. a., Debets, A.J.M., van der Oost, J., Hoekstra, R.F., Tawfik, D.S. & de Visser, J.A.G.M., 2011. Initial mutations direct alternative pathways of protein evolution. *PLoS Genetics*, 7(3), pp.1–11.

Sanjuán, R. & Elena, S.F., 2006. Epistasis correlates to genomic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(39), pp.14402–14405.

Sanjuán, R., Moya, A. & Elena, S.F., 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*, 101(43), pp.15376–15379.

Schenk, M.F., Szendro, I.G., Salverda, M.L.M., Krug, J. & De Visser, J.A.G.M., 2013. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Molecular Biology and Evolution*, 30(8), pp.1779–1787.

Schleif, R., 2000. Regulation of the L-arabinose operon of Escherichia coli. *Trends in Genetics*, 16(12), pp.559–565.

Schleif, R., 2003. AraC protein: a love-hate relationship. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 25(3), pp.274–82.

Schleif, R.F., 1993. *Genetics and Molecular Biology*, Johns Hopkins University Press.

Silander, O.K., Tenaillon, O. & Chao, L., 2007. Understanding the evolutionary fate of finite populations: The dynamics of mutational effects. *PLoS Biology*, 5(4), pp.922–931.

Stent, G.S., 1964. The Operon: On Its Third Anniversary. *Science*, 144 (3620), pp.816–820.

Tjian, R. & Maniatis, T., 1994. Transcriptional activation: A complex puzzle with few easy pieces. *Cell*, 77(1), pp.5–8.

Wade, M.J., Winther, R.G., Agrawal, A.F. & Goodnight, C.J., 2001. Alternative definitions of epistasis: Dependence and interaction. *Trends in Ecology and Evolution*, 16(9), pp.498–504.

Wang, Y., Diaz Arenas, C., Stoebel, D.M. & Cooper, T.F., 2012. Genetic background affects epistatic interactions between two beneficial mutations. *Biology Letters*, pp.1–4.

Weinreich, D.M., Watson, R. a & Chao, L., 2005. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, 59(6), pp.1165–1174.

Wittkopp, P.J. & Kalay, G., 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1), pp.59–69.

Wittkopp, P.J., 2005. Genomic sources of regulatory variation in cis and in trans. *Cellular and Molecular Life Sciences*, 62(16), pp.1779–1783.

Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews. Genetics*, 8(3), pp.206–16.

You, L. & Yin, J., 2002. Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage T7. *Genetics*, 160(4), pp.1273–1281.

Appendix

Protocol for Preparation of Electrocompetent Cells

Day 1

1) Autoclave 1L ddH_2O and 100 mL of 10% glycerol, 100 mL of LB, 300 mL or 1L sized baffled flask. 100 mL of culture yields about 14 aliquots of electrocompetent cells. Culture volume to flask ratio should be around 1:10 or 1:5.
Following instructions are for 100 mL of culture.

2) Inoculate 2 mL of LB with cells, grow overnight, shaking at 37 °C in LB.

Day 2

1) Pre-chill centrifuge, rotors, sterile water, 10% glycerol, cuvettes, pipettes (25 and 10 mL), 50 mL Falcon tubes, and sterile Eppendorf tubes at 4 °C.

2) Inoculate 100 mL LB in flask with 1 mL of overnight culture (1:100 dilution, but can also use 1:50 dilution ratio; use flask with baffler). Measure OD_{600} of 1 mL of LB, to use as spectrophotometer blank.

3) Incubate culture, shaking at 37 °C, until it reaches OD_{600} of 0.5.

4) Immediately take flask from incubator and submerse it in ice-water slurry. Swirl flask for a few minutes. Leave flask in slurry for at least 10 min. Quick cooling of culture is crucial. Culture should be kept cold for the rest of the protocol.

5) Transfer culture to a pre-chilled 50 mL Falcon tube.

6) Centrifuge for 10 min at 3000 x g, discard supernatant and re-suspend pellet in 35 mL of ice-cold ddH_2O.

7) Incubate on ice for 5 min.

8) Centrifuge for 10 min at 3000 x g, discard supernatant and re-suspend pellet in 35 mL of ice-cold ddH_2O.

9) Incubate on ice for 5 min.

10) Centrifuge for 10 min at 3000x g. Discard supernatant and re- suspend pellet in 1.5 mL of ice-cold 10% glycerol. Transfer to 2 mL Eppendorf tube.

11) Incubate on ice for 5 min.

12) Centrifuge for 2 min at top speed on tabletop centrifuge. Discard supernatant and re- suspend pellet in 1mL (final volume) ice-cold 10% glycerol.

13) Use cells directly (keeping them on ice) or freeze aliquots of 50 µL at -80 °C. If freezing cells, it is best to flash freeze in liquid nitrogen, but not absolutely necessary.